

# 比較記述テキスト生成のための属性選択

飯田 諒<sup>1,a)</sup> 飯田 龍<sup>1,b)</sup> 徳永 健伸<sup>1,c)</sup>

**概要：**言語生成における参照表現生成では、特定の対象を曖昧性なく簡潔に指示する言語表現を生成することを目的として研究が行われてきた。参照表現生成が対象の「弁別」を重視するのに対し、我々は、複数の対象を横並びで比較するのに十分な情報を読み手に伝えることを目的としたテキスト（比較記述テキスト）の生成を目指している。このための基礎データとして、複数の対象を複数の観点から評価した評価表から人間が生成した比較記述テキストを収集し、比較記述テキストの生成における対象の属性選択の傾向について分析を行った。本稿では比較記述テキストの収集方法、テキスト中で言及された対象とその属性のアノテーション結果、さらに属性選択の分析結果について報告する。

## 1. はじめに

計算機によるテキストの生成は自然言語処理の研究の一分野として研究されてきた [10], [12]。とりわけ特定の対象を曖昧性なく簡潔な表現で他の対象と区別する参照表現の生成はテキスト生成の中心的な課題として扱われてきた [5]。参照表現生成における主な関心は、指示対象を他の対象から弁別するために、指示対象が持つ様々な属性の中からどのような属性を選び、言語化するかという点にある。このために種々の属性選択のアルゴリズムが提案されている [1], [2]。これに対して我々は複数の対象が与えられた時に、その中から聞き手が各自の規準にしたがって対象を選択することを想定し、対象を選択するために各対象についての十分な情報を含むテキストを生成することを目指している。したがって、このようなテキストでは、聞き手の嗜好やシステムの嗜好を前提とせず、聞き手が対象を比較する上で必要な情報を、すべての対象に対して中立的に記述することになる。このようなテキストを以下、**比較記述テキスト**と呼ぶ。対象の特徴を記述するという点では、比較記述テキストも参照表現と同じであるが、その目的が他の対象からの「弁別」ではなく、他の対象との「比較」であるという点が異なっている。参照表現生成のための属性選択アルゴリズムを各対象について個別に適用すれば、「弁別」という観点からは十分な情報が記述できるが、これがそのまま対象を「比較」する上で有用な情報となりうるかどうかは明らかではない。

比較記述テキストの利用例としては、インターネット上の商品の評判サイトなどに見られる商品の比較情報などをテキスト化し、要約することなどが考えられる。商品比較

は表 1 に示すような表の形で提示されることが多い。このような表は各対象の各属性について評価値を付与した一種の数値データであると考えられる。一見、このような表をテキストとして言語化することは、「要約」という観点からは逆行しているように見えるかもしれない。しかし、実際に人間がこの表から情報を読み取るには、表の内容を解釈し、それぞれの商品が他の商品に比べてどのような特徴があるかを把握する必要がある。このように数値データを入力として、データを解釈し、テキストを生成する研究はこれまでも行われてきた。例えば、気象情報データから天気予報士が読み上げる天気予報の原稿を生成するシステム [11] や株価の変動データをテキスト化するシステム [6] などが構築されている。本研究で対象とする生成処理は、数値データからテキストを生成するという点では既存研究と同じであるが、既存研究は個別の場所の天気や株価について独立にテキストを生成するのに対し、本研究では対象横断的に中立的なテキストを生成する点が大きく異なる。

また、対象を説明する際に、それと類似した物と比較し、相違点を説明することでわかりやすい説明ができると言われている [4], [7]。このように、他の対象と対比することで、言及したい対象を説明する方略は初期のテキスト生成でも使われてきた [8]。例えば、Direct comparison [9] という比較手法では、2つの対象を比較して、類似点と相違点を述べる戦略をとる。この際に、2つの対象をそれぞれ偏りなく記述するという点では本研究の比較記述テキストの目的と類似しているが、Direct comparison は対象が2つに限定されているため、3つ以上の対象を比較する場合の生成の戦略については言及されていない。また他の比較記述に関する研究では、注目された記述対象が他の対象とどう異なるかという比較を対象にしており [9]、比較対象すべてに対して中立的な立場でテキストを生成することについては考慮されていない。

これらの背景を踏まえて、本稿では比較記述テキスト生成の基礎データを構築するために行ったテキスト収集実験

<sup>1</sup> 東京工業大学  
Tokyo Institute of Technology

<sup>a)</sup> iida.r.ab@m.titech.ac.jp

<sup>b)</sup> ryu-i@cl.cs.titech.ac.jp

<sup>c)</sup> take@cl.cs.titech.ac.jp

表1 実験参加者に与えられる評価表の例

属性	テレビ A	テレビ B	テレビ C	テレビ D	テレビ E	属性の説明
デザイン	★★★★★	★★	★★	★	★★★	見た目のよさ, 質感
サイズ	★★	★★	★	★★	★	省スペース性・コンパクトさ
機能性	★	★★★★★	★★★★	★★	★★	機能が充実しているか
応答性能	★★★	★	★★★★★	★★★★	★★	動画などで残像が残らないか
音質	★★	★★★	★★★	★★★★★	★	音質の良さ
画質	★★★	★	★★	★★	★★★★★	映像の鮮明さなど
操作性	★★	★★★	★★★★★	★★★★	★★	リモコンや本体の操作のしやすさ
価格	75,000	70,000	80,000	75,000	75,000	価格
満足度	★★★	★★★	★★★	★★★	★★	総合的に評価した値

の詳細, および, 収集したテキスト中で各対象のどの属性が言及されたかをアノテーションした結果について述べる。さらに, このアノテーションした結果を利用して, 実際のテキスト生成を行う際に与えられた表のどこを言及するかの傾向を調査した結果についても報告する。以降では, この表のどの箇所を言及するかを選択する問題を**属性選択**と呼ぶ。まず, 比較記述テキストコーパスの構築方法とその結果を2節で示し, 提案する比較記述テキスト生成モデルの全体像を3節で説明する。4節で属性選択モデルを提案し, その評価結果を5節で報告する。最後に, 6節でまとめと今後の課題について述べる。

## 2. 比較記述テキストコーパスの構築

比較記述テキストコーパスを構築するにあたり, テキストの収集実験および, 属性選択に関する情報をアノテーションした。以下で, 収集実験の詳細, アノテーションの内容, アノテーションされた情報の分析結果について述べる。

### 2.1 比較記述テキストの収集実験

比較記述テキストを生成するために, 表1に示すような10種類の表(以下, この表を**評価表**と呼ぶ)を用意し, 20名の記述者に各評価表について各製品の特徴を比較して述べるテキストを記述させ, 合計200の比較記述テキストを収集した。本研究で扱う比較の対象として, 特別な背景知識を必要とせずに扱える領域として家電製品を選択した。各家電製品に対する評価項目(属性)は価格.com<sup>\*1</sup>のレビューに利用されている項目を用い, 比較する各製品の各項目について表1に示すような5段階評価を付与した評価表を用意した。各属性の値は1から5の5段階評価だが, 例外的に属性「価格」の値のみ価格の数値を用いた(以下, これらの値を**評価値**と呼ぶ)。属性名は価格.comで使用されているものをそのまま利用しているが, 属性の評価値は異なる表の特性を調べるために, 後述するいくつかの要因を考慮して恣意的に決定した。本研究では, 5段階の評価を相対評価, 「価格」のような数値をとる評価を絶対評価と呼ぶ。テレビ, 掃除機, 電子レンジ, DVDレコーダ, 炊飯器の5種類を製品クラスとして選択し, 各製品クラスに対してそれぞれ2つの評価表を作成し, 合計10種の評価表

を用意した。また, 評価表の作成時には, 以下の5つの要因を考慮した。

#### 製品数

比較する製品の個数によって比較記述に違いがでるか調べるために, 表中の製品の個数を変化させてテキストを記述させた。具体的な表の個数は製品が5個の場合が3つ, 6個の場合が3つ, 7個の場合が4つである。

#### 評価値の偏り

評価表中の相対評価の各セルの評価値は1から5個の星の数で表現されている。特定の評価値を特に際立たせるために, 星4つあるいは星2つの評価を避けて表を作成した。これにより星5つや星1つの評価値の顕現性が高まることが期待できる。星4つを含まない表を2つ, 星4つと星2つを含まない表を2つ作成した。

#### 各製品に特有な属性の有無

各製品が他の製品と異なる特有の属性の評価値を持つかどうかで2通りのパターンを考えた。例えば, 表1において各テレビはそれぞれ異なる属性で唯一の星5つの評価値を持つ。このような状況で参照表現生成のアルゴリズムを適用すると各テレビに特有の属性だけを言及することになる。この要因を導入することで, 比較記述で人間が参照表現生成とは異なる方略で比較記述テキストを生成することを確かめる。製品ごとに特有の評価値を含む表を4つ, 含まない表を6つ作成した。

#### 属性間の相関の有無

属性間に相関関係がある場合に生成されるテキストが異なるかを調査する。ここで, 属性間の相関とは, 例えば, 画質の評価値が上昇するとそれに伴って価格が上昇することを表す。このような相関がある表を5つ, 相関がない表を5つ作成した。

#### 属性間の相関における反例の有無

属性間の相関関係について一部相関しない場合を考える。例えば, 製品AとEの間では画質が高い場合に価格が高くなっている一方, 製品Fは画質が高いが極端に価格が安くなっているという場合を考える。前述の「属性間の相関の有無」で作成した相関のある表5つに関して, このような反例を含む場合を3つ, 含まない場合を2つ作成した。

\*1 <http://kakaku.com>

表2 各評価表と要因の関係

表の種類	属性数	製品数	評価値の偏り	特有な属性の有無	相関の有無	相関の反例
テレビ 1	9	5	○	○	×	×
テレビ 2	9	7	×	×	○	○
掃除機 1	9	5	×	×	×	×
掃除機 2	9	7	○	○	○	×
電子レンジ 1	9	5	×	○	×	×
電子レンジ 2	9	7	○	×	○	×
DVD レコーダ 1	10	6	○	×	×	×
DVD レコーダ 2	10	6	×	×	○	○
炊飯器 1	8	6	×	×	×	×
炊飯器 2	8	7	×	○	○	○

表2に各評価表と5つの要因の関係を示す。実験参加者にはこの10個の評価表に関し、以下の指示を与えてテキストを作成させた。

- 一つの評価表について生成するテキストは400字程度とする。
- 特定の製品を薦めるのではなく、中立的な立場ですべての製品について記述する。
- 特徴ごとに記述するのではなく、製品ごとにその特徴を記述する。
- 満足度と価格だけで各製品を説明しない。
- 複数の製品をまとめて記述してもよい。

表1のテレビの評価表に対して1人の記述者が書いたテキストを例として示す。

テレビAは他の製品に比べて圧倒的にデザイン性が優れているが、機能性は劣っている。映像や音はごく標準で、価格も平均的だ。テレビBは充実した機能を持つ反面、残念な画質だ。音やリモコンの操作性に対する評価は悪くなく、価格はお手頃といえるだろう。テレビCは動画が良く、リモコンが非常に使いやすい。しかし映像は粗めで本体サイズが大きい。そして他と比べて価格はやや高めだ。テレビDは音質が素晴らしい。だが映像への評価は普通。見た目の悪い点がマイナスポイントとしてあげられている。テレビEは映像が格段に鮮明で美しい。しかし、音に関してはいま一つだ。その他の機能評価も平均よりやや低い。テレビ本体の見た目もなかなか良いが、ユーザーの総合満足度は低めだ。各製品とも特徴は様々だが、テレビA、B、C、Dに対する総合的な満足度はほぼ同水準となっている。ただし、テレビEだけがやや低い結果となった。

## 2.2 選択された属性のアノテーション

本研究で構築するコーパスは比較記述における人間の属性選択の傾向を明らかにすることを目的としている。そのため、2.1で説明した実験を通じて収集したテキストに対し、テキスト中のどの部分がどの対象のどの属性に言及しているかをアノテーションした。アノテーションの手順としては、まずテキスト中の属性への言及している表現を抽出し、その表現に対して「選択された属性」、「言及してい

る製品名」、および「評価の極性」を情報として付与する。ここで、属性への言及に対して極性を付与するのは、各製品の記述の中で肯定表現と否定表現がどのような割合で含まれているかを調査するためである。例えば、「テレビAはデザインが、特に優れていますが、機能が少ないです」という記述に対しては、「デザインが、特に優れています」が属性への言及であるため、この表現に対して、製品名が「テレビA」、属性が「デザイン」、極性が「肯定」という情報を付与する。

アノテーションには、アノテーションツール Slate [3] を使用した。また、ある対象がテキスト中で言及されているかの判断が困難な場合があるため、そのような場合にどのようにアノテーションするかを決定する必要がある。そこで、我々が既存研究 [13] で行ったように、下記の手順で判断が揺れる事例を検出し、それに対して個別にアノテーションの基準を策定することで、判断の揺れを解消した。

- (1) 評価表とその表から生成されたテキストに対し、2人がアノテーションを行う。
- (2) アノテーション結果の不一致を検出し、その不一致の原因を分析することで、アノテーションの基準を更新する。

アノテーションの判断が揺れる典型的な例として間接的に対象が触れられている場合がある。例えば、「レコーダA、C、Fは同一価格ですが、Fの満足度が最も高くなっています」といった記述では、レコーダFの満足度に加えて、間接的にはレコーダA、Cの満足度にも言及されているとみなすこともできる。このような場合、どのような記述の場合に間接的な言及であるかの判断が困難であるため、まずは作業の揺れを無くすためにアノテーションの基準として「間接的に対比されている情報はアノテーションしない」という基準を採用した。これにより、上の例の場合はレコーダFの満足度のみがアノテーションされる。本研究では与えられた数値データから最終的には対象間の「比較」を適切に捉えることを目的としているため、今後の課題として、揺れが生じた間接的に対象が言及される場合についてもなんらかのアノテーション基準を策定し、網羅的な比較の言及についてアノテーションを行うことを考えている。

表5 評価表（テレビ1）の製品-属性の対の各言及率

属性	テレビ A	テレビ B	テレビ C	テレビ D	テレビ E	属性の説明
デザイン	1.00	0.20	0.05	0.90	0.20	見た目のよさ、質感
サイズ	0.10	0.15	0.90	0.10	0.70	省スペース性・コンパクトさ
機能性	0.90	1.00	0.15	0.30	0.15	機能が充実しているか
応答性能	0.15	0.85	0.95	0.30	0.15	動画などで残像が残らないか
音質	0.35	0.15	0.05	0.95	0.80	音質の良さ
画質	0.35	0.95	0.30	0.35	1.00	映像の鮮明さなど
操作性	0.40	0.20	1.00	0.35	0.15	リモコンや本体の操作のしやすさ
価格	0.30	0.95	0.95	0.40	0.20	価格
満足度	0.45	0.35	0.45	0.30	0.70	総合的に評価した値

表3 表ごとの属性選択

表の種類	総数	平均	最大	最小
テレビ 1	45	21.6	33	12
テレビ 2	63	30.1	48	21
掃除機 1	45	24.3	35	13
掃除機 2	63	24.4	43	12
電子レンジ 1	45	24.7	35	15
電子レンジ 2	63	28.9	44	15
DVD レコーダ 1	60	27.0	41	14
DVD レコーダ 2	60	25.5	37	10
炊飯器 1	48	25.7	37	13
炊飯器 2	56	27.4	35	22

表4 各記述者の属性選択数の割合

記述者	平均	記述者	平均
d1	0.665	d11	0.391
d2	0.416	d12	0.402
d3	0.558	d13	0.572
d4	0.521	d14	0.472
d5	0.638	d15	0.579
d6	0.466	d16	0.407
d7	0.496	d17	0.337
d8	0.339	d18	0.424
d9	0.402	d19	0.478
d10	0.524	d20	0.488
		平均	0.479
		標準偏差	0.0896

### 2.3 アノテーション結果の分析

次に、属性選択の傾向分析のための予備調査として、構築したコーパスをいくつかの観点から分析した。まず、評価表ごとに属性選択数がどう異なるかを調査した。この結果を表3に示す。表3の「平均」は各表における記述者が言及した属性数の平均を表し、「最大」（「最小」）は各表における記述者が言及した属性数の最大数（最小数）を表している。表3から評価表ごとに属性選択数が異なることがわかる。また、どの評価表においても、選択された属性の最大数と最小数に大きな差があり、記述者によって属性選択数が大きく異なることがわかる。さらに、記述者間で言及する属性にどのくらい違いがあるかについても調査した結果を表4に示す。表4の「平均」は各記述者が各表で

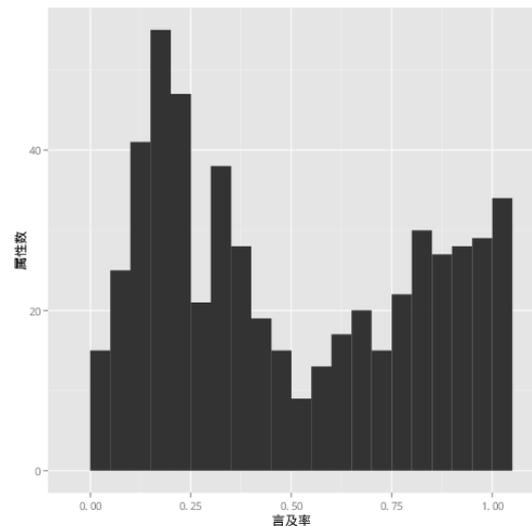


図1 言及率の分布

言及した属性数の全属性数に対する割合の平均を表している。属性選択の割合の平均については、最小値（記述者d17の平均値）と最大値（記述者d1の平均値）を比較すると約0.33もの差があり、記述者ごとに言及する属性に大きな違いがあることがわかる。さらに、全記述者のうち製品-属性の対を何人が記述したかの割合（以下、言及率）の分布をテレビ1（表1）について調査した結果を表5に示す。表5より、ほとんどの記述者が言及する属性、逆にほとんどの記述者が言及しない属性がある一方で、半数が選択する属性も存在していることがわかる。

さらに言及率について調査するために、10種の評価表の全属性に対する言及率をヒストグラムとして出力したものを図1に示す。横軸が言及率、縦軸はその言及率を持つ属性の個数を示している。この図より、言及率が0から0.15までの区間で属性数は単調増加しているものの、0.20以降で減少し、0.50から再度増加していることがわかる。この言及率の分布から、言及するか否かの判断が偏る属性が多いことがわかる。ただし、言及率が半数程度の属性も少ないわけではないため、人間が比較をする上でなぜこのような違いが存在するのかをさらに調査する必要がある。

この収集したコーパスに基づき、比較記述テキスト生成モデルを構築することを考えるが、図1で見たように、人間でもさまざまな生成の戦略をとり、その結果、言及する

対象-属性の対が異なるため、単純な二値分類の問題として言及するカラムを分類することは困難である。そこで、3節ではこの言及のばらつきを考慮した生成モデルについて提案する。

### 3. 比較記述テキスト生成モデル

テキスト生成のモデルは一般に、内容の決定と構成を行うテキスト・プランニング、構成素の集約、語選択、参照表現の生成などを行うマイクロ・プランニング、文法的に正しい文字列を出力する表層化の3つの部分処理から成るとされている [10]。我々も基本的にこのモデルにしたがう。このうち本稿で扱うのは主に内容の決定の処理であり、特に個々の対象を記述する際に言及すべき対象の属性を選択する属性選択の問題を扱う。これは参照表現生成では中心的な課題として研究されてきた問題である。参照表現生成の場合は、他の対象との弁別が重要となるため、曖昧性なく弁別するための属性の集合を規定しやすい。しかしながら、図1の例からもわかるように、比較記述テキストにおいては各対象のどの属性に言及するかは個人差が大きく、参照表現のように明確な属性集合を規定することが困難である。

また、参照表現生成では、指示対象を表す名詞句を生成することが多いが、比較記述テキストの場合は、複数の対象を記述する節や文の生成を考慮しなければならないため、内容の構成の方略や統語的な制約によっては、各対象で選択された属性が表層化されるとは限らないし、逆に属性選択の段階で選択されなかった属性も表層化しなければならない可能性もある。例として以下のような記述を考えよう。

テレビ A は、デザイン性においてはもっとも高評価で、それはテレビ D と比べて遥かに優れていますが、機能性においてはもっとも低評価です。

この対比を使った記述では、テレビ A とテレビ D のデザイン性について言及しているが、属性選択の段階ではテレビ A とテレビ D の両方のデザイン性が選択されているとは限らない。仮にテレビ D のデザイン性が選択されていなくても、テレビ A のデザイン性を強調するためには、対比方略を使う上でテレビ D のデザイン性も選択しなければならない。このように属性選択の課題は必ずしもそれ自身が独立して考えられるものではなく、その結果は内容の構成や統語的な構成、語選択などの後段の処理の影響も受ける可能性がある。以上のような理由から本研究における属性選択の位置付けは、各対象の属性集合を決定するのではなく、後段の処理に対する柔軟性を残すために、各対象の属性の選択されやすさを定量的に表すことを目的とする。

### 4. 属性選択の回帰モデルとその素性

我々は現在3節で述べた生成モデルにおける対象の属性選択に取り組んでおり、各対象の属性に対して一種の言及されやすさを定量的に測定したものを与えることを中心に分析を行っている。言及されやすさのスコア付けを目的と

して、本研究では回帰モデルを利用する。この理由としては、回帰モデルを利用することで、各属性の言及されやすさを数値化することができるためである。また、数値化を行い、それを属性選択に関するスコアとすることで、今後の課題として考えている対象間もしくは属性間の依存関係を考慮した属性選択（「対比」方略など）の実現や、表現の順序を考慮した言語生成を行う上でも、柔軟なテキスト生成を可能にすると考えられる。回帰モデルでは言及率を目的変数として設定し、後述する種々の特徴量を説明変数として設定する。特徴量を考える上では、製品を横並びで比較する上で特徴的な評価値となるのは評価値の平均から離れているものであるという考えに基づき、平均値からの偏差を特徴量とした。以下、どのような範囲で評価値の平均を取るかで属性内偏差、製品内偏差、表内偏差の三種類の偏差を考える。これらの偏差は、それぞれ、同一属性内、同一製品内、表全体で評価値の平均 ( $\mu$ ) と標準偏差 ( $\sigma$ ) を計算し、各評価値 ( $v$ ) に対して以下の式で与えられる  $dv$  として定義される。

$$dv = \frac{|v - \mu|}{\sigma} (\text{属性内・製品内}), \quad dv = |v - \mu| (\text{表全体})$$

これらの違いは偏差の規準となる平均をどの範囲で計算するかである。

また、特徴量を計算する上で、絶対評価と相対評価を同等に扱えない場合がある。そこで正規化を行い、すべての評価値を0から1の間の数値に変換する。正規化前と正規化後の評価表の例を表6に示す。この正規化後の評価表を用いて以下に示すそれぞれの特徴量を計算する。

表7 特徴量の一覧

特徴量	言及率との相関係数
属性内偏差	0.767
製品内偏差	0.727
表内偏差	0.824

回帰分析を行う上で目的変数と説明変数の相関係数の程度は、回帰分析を行うことが妥当か否かの判断の1つの指標となる。そのため、言及率と各特徴量の相関係数を表7に示す。表7が示すように、言及率と今回用意した三つの偏差の相関係数はいずれも高いことがわかる。このようにして計算される特徴量を説明変数、言及率を目的変数として回帰分析を行い、属性の言及されやすさを推測するためのモデルを導出した。

### 5. 回帰モデルの評価

4節で導入した回帰モデルの評価を複数の観点から行った。以下でそれぞれの評価方法および評価結果について述べる。

#### 5.1 回帰モデルの当てはまり評価

用意した特徴量と言及率を用いて回帰分析を実施した。表8に各特徴量を説明変数として実施した回帰分析の結果を示す。回帰分析の精度、つまり回帰モデルの当てはまり

表6 正規化する前と後の評価表

正規化前					正規化後				
特徴	製品 A	製品 B	製品 C	製品 D	特徴	製品 A	製品 B	製品 C	製品 D
属性 A	★★★★★	★★★	★★★	★	属性 A	1.0	0.5	0.5	0
属性 B	★★★	★★	★★	★★	属性 B	0.5	0.25	0.25	0.25
属性 C	★★	★★★★	★★★★★	★★★★★	属性 C	0.25	0.75	1.0	0.75
属性 D	1,000	2,000	2,000	3,000	属性 D	0	0.5	0.5	1.0

表8 回帰分析の結果

説明変数	決定係数
属性内偏差	0.579
製品内偏差	0.527
表内偏差	0.678
属性内偏差 + 製品内偏差	0.657
属性内偏差 + 表内偏差	0.686
製品内偏差 + 表内偏差	0.688
属性内偏差 + 製品内偏差 + 表内偏差	0.696

表10 交差検定の結果

テストデータ	平均平方二乗誤差	決定係数
テレビ 1	0.152	0.794
テレビ 2	0.209	0.588
掃除機 1	0.261	0.330
掃除機 2	0.141	0.800
電子レンジ 1	0.148	0.799
電子レンジ 2	0.172	0.695
DVD レコーダ 1	0.206	0.661
DVD レコーダ 2	0.182	0.584
炊飯器 1	0.185	0.532
炊飯器 2	0.160	0.772

を測る指標として自由度調整済み決定係数を利用した。

この結果から、最も当てはまりのよい特徴量は表内偏差であることがわかる。そこで、5.2以降では表内偏差を特徴量とした回帰モデルを利用する。また、複数の説明変数を用いた重回帰分析も同様に実施した。表7の各特徴量を組み合わせた重回帰分析の結果を表8に示す。この表が示すように、各特徴量を組み合わせることで決定係数が上昇することがわかった。ここで、新たな特徴量として、評価値の平均より大きい小さいかの二値で表す特徴量を導入した。この特徴量と単回帰分析において最も結果の良い特徴量である表内偏差を2つの説明変数として回帰分析したところ、回帰分析の精度が向上し、自由度調整済み決定係数が0.712となった。この理由としては、記述者は悪い評価より良い評価を好んで選択する傾向があるからだと考えられる。表9は相対評価の各評価値における言及率の平均を示している。この表が示す通り、評価値として良い評価の方が言及がされやすく、人は悪い評価の属性よりも良い評価の属性を用いて、テキストを生成する傾向にあることがわかった。

表9 各評価値の言及率の平均

評価値	1	2	3	4	5
言及率の平均	0.80	0.27	0.18	0.67	0.96

## 5.2 交差検定による評価

5.1では今回用意した評価表の種類に関係なく、すべての評価表を統合したデータに対して回帰モデルの当てはまりを評価した。そのため、各評価表の間でどの程度同じ回帰モデルが当てはまるかについて考慮しておらず、各評価表から導出される回帰モデルの共通性を扱っていなかった。そこで1つの評価表をテストデータ、その他9つの評価表を訓練データとした交差検定を実施した。表10に交差検定の結果を示す。

表10に示した結果より、「掃除機1」の決定係数が低く、

訓練データから導出した回帰モデルでは当てはまりが悪いことがわかった。この原因としては、「掃除機1」の表における評価値の平均が他の表における評価値の平均よりも高いことが考えられる。平均が高いことによって、必然的に良い評価への偏差が小さく、悪い評価への偏差が大きくなるため、人の属性選択の傾向「良い評価を悪い評価よりも選ぶ」ことに反してしまうからである。

## 5.3 属性選択の再現性評価

回帰モデルで求めた各製品における属性の言及されやすさのスコアが与えられたときに、そのスコアの順位に応じて言及するか否かに分類するモデルを考える。このモデルを採用した場合、すでに収集済みの比較記述テキスト中の属性選択をどのくらい再現できるかを調査する。比較記述テキストは各対象を説明するテキストであるため、各製品ごとにどの属性について言及するかの問題に着目し、まず各記述者が各製品に対してどれくらいの属性数を利用してテキストを生成していたかを調査した。また、各属性の言及率の高い属性から、前述した属性選択数の平均（切り上げ）まで選出し、それらを言及とした場合に正解という設定を考える。例えば、表11の各セルは製品-属性の対の言及率であり、一番下の行が各製品を説明する際に記述者が選択した属性数の平均を表している。このような場合、製品Aから製品Cまでは2つの属性を言及率が高いものから選択して、それらの属性を正解として選択した。製品Aであれば、属性Aと属性Dが選択すべき属性となる。

この問題に対して、4節で導入した回帰モデルが言及すべき属性をどの程度正しく当てることができるかを考える。回帰モデルにおいても、各製品について属性をいくつまで選択するかという判断基準は同じだが、回帰モデルで

表 11 問題設定の例

特徴	製品 A	製品 B	製品 C	製品 D
属性 A	0.56	0.40	0.47	0.16
属性 B	0.11	0.30	0.40	0.40
属性 C	0.08	0.25	0.13	0.04
属性 D	0.25	0.05	0.00	0.40
属性選択数の平均	1.8	2.0	1.5	2.5

計算される言及率の大きいものから選択する。属性選択の再現評価の結果を表 12 に示す。横が記述者の選択から決定した正解、縦が回帰モデルが出力した結果を表す。結果として、収集したテキストから選択すべきと判断した属性の再現率は  $237/(237+52) = 0.820$  であり、高い割合で選択すべき属性をカバーできることがわかった。また出力全体の精度は、 $(237+207)/(237+52+52+207) = 0.810$  であり、高い精度で選択すべきか否かを当てることができていることがわかった。

表 12 属性選択の再現度の評価結果

モデル \ 正解	言及する	言及しない
言及する	237	52
言及しない	52	207

## 6. おわりに

本稿では、比較記述テキスト生成のための基礎データとして我々が構築した比較記述コーパスの概要とテキスト生成における対象の属性選択の傾向分析について述べた。まずコーパス構築のためのテキスト収集実験を実施し、20名の記述者に10種類の家電製品の評価表から比較記述テキストを生成させ、200の比較記述テキストを収集した。次に収集したテキストに対して、記述者が言及した製品の属性に関する選択を明示的にするためのアノテーションを行い、アノテーション結果をもとに、比較記述テキストに見られる特徴について述べた。さらに、テキスト生成の第一段階として重要な属性選択の傾向分析を行い、各属性の言及されやすさを推測するための回帰モデルの分析・評価を行った。言及されやすさを目的変数とした回帰分析では、説明変数として表全体から計算した評価値の平均からの差が最も有効であることがわかった。また、この特徴量を利用した回帰分析から得られた属性の言及されやすさを数値化したモデルでは、実際に人間が行った属性選択を0.820の割合で再現することができた。

さらに、評価表に共通の回帰モデルが有効であるかどうかの評価を行った結果、当てはまりの悪い評価表も存在することがわかった。このように当てはまりの評価表の特徴を分析し、各評価表ごとに適した特徴量を利用可能なモデルを開発する必要がある。さらに属性間の依存関係のアノテーションを行い、その分析を行うことも重要である。人間の記述に見られる比較方略を収集したテキストから明らかにし、実際に自動生成する際にその比較方略を再現することが望ましい。また、新たなテキスト収集も今後の課

題として考えている。新たなテキスト収集では、今回収集したテキストを通じて調査すべきだとわかったことを考慮する。例えば、価格のような絶対的な値で評価された属性と5段階の星の数を用いて相対的な値で評価された属性とでは、属性選択の間で差があるかなどを調査していきたい。

## 参考文献

- [1] Dale, R.: Cooking up referring expressions, *Proceedings of the 27th Annual Meeting of the Association for Computational Linguistics*, pp. 68–75 (1989).
- [2] Dale, R. and Reiter, E.: Computational interpretation of the Gricean maxims in the generation of referring expressions, *Cognitive Science*, Vol. 19, No. 2, pp. 233–263 (1995).
- [3] Kaplan, D., Iida, R., Nishina, K. and Tokunaga, T.: Slate – A tool for creating and maintaining annotated corpora, *Journal for Language Technology and Computational Linguistics*, Vol. 26, No. 2, pp. 89–101 (2012).
- [4] Karasimos, A. and Isard, A.: Multi-lingual evaluation of a natural language generation system, *Proceedings of 4th International Conference on Language Resources and Evaluation (LREC 2004)*, pp. 829–832 (2004).
- [5] Krahmer, E. and van Deemter, K.: Computational generation of referring expressions: A survey, *Computational Linguistics*, Vol. 38, No. 1, pp. 173–218 (2012).
- [6] Kukich, K.: Design of a knowledge-based report generator, *Proceedings of the 21st Annual Meeting of the Association for Computational Linguistics (ACL 1983)*, pp. 145–150 (1983).
- [7] Marge, M., Isard, A. and Moore, J.: Creation of a new domain and evaluation of comparison generation in a natural language generation system, *Proceedings of the 5th International Natural Language Generation Conference (INLG 2008)C*, pp. 169–172 (2008).
- [8] McKeown, K.: *Text Generation*, Cambridge University Press (1985).
- [9] Milosavljevic, M. and Dale, R.: Strategies for comparison in encyclopedia descriptions, *Proceedings of 8th International Natural Language Generation Workshop (INLG 1996)*, pp. 161–170 (1996).
- [10] Reiter, E. and Dale, R.: Building applied natural language generation systems, *Natural Language Engineering*, Vol. 3, No. 1, pp. 57–87 (online), DOI: doi:10.1017/S1351324997001502 (1997).
- [11] Sripada, S. G., Reiter, E., Hunter, J. and Yu, J.: Exploiting a parallel TEXT-DATA corpus, *Proceedings of the Corpus Linguistics 2003 conference*, pp. 734–743 (2003).
- [12] Uszkoreit, H.: Language generation, *Survey of the State of the Art in Human Language Technology*, Cambridge University Press, chapter 4 (2010).
- [13] 飯田 諒, 飯田 龍, 徳永健伸: 製品の比較記述コーパスの構築, 言語処理学会第20回年次大会発表論文集, pp. 912–915 (2014).